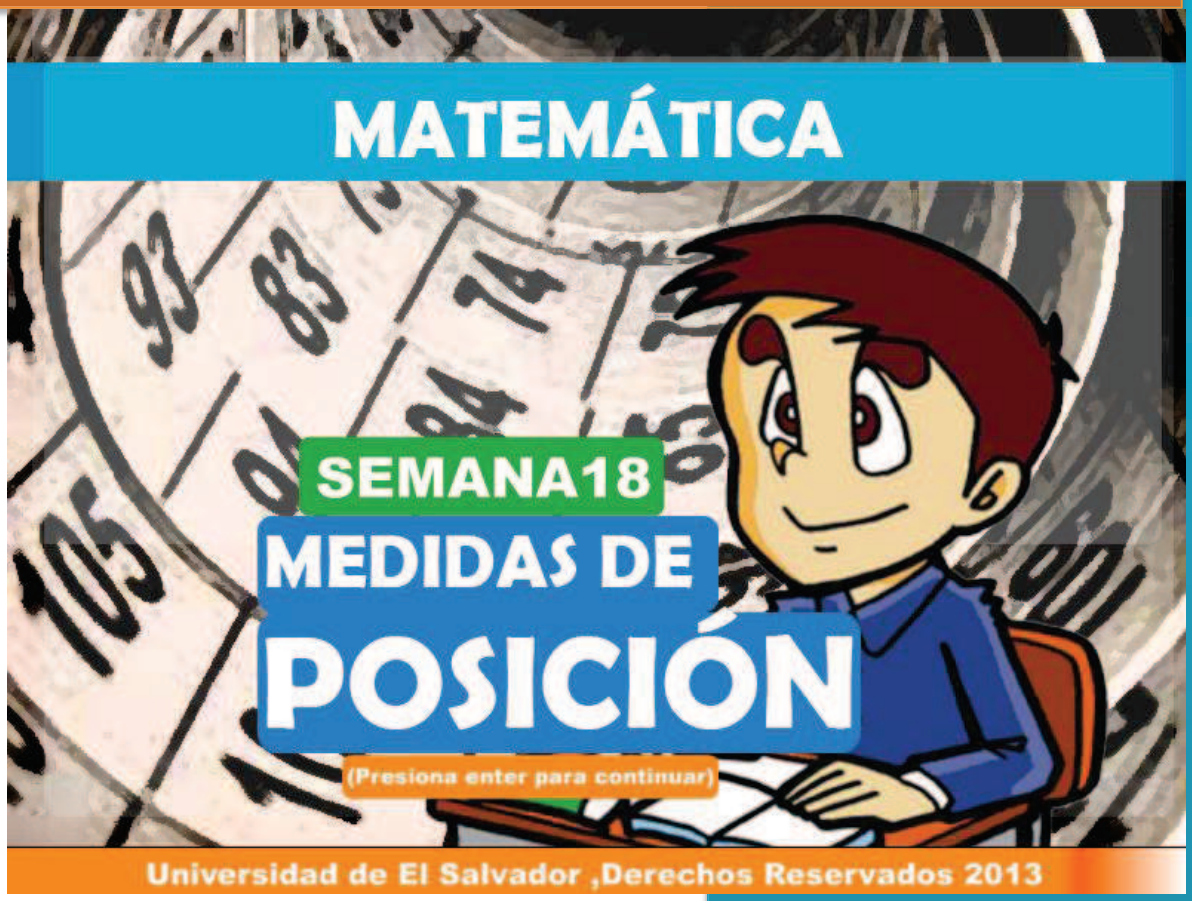
UNIVERSIDAD DE EL SALVADOR
Curso de Nivelación 2014

MATEMÁTICA



Universidad de El Salvador, Derechos Reservados

CURSO DE MATEMÁTICA EN LÍNEA

Contenido

MEDIDAS DE POSICIÓN 1
MEDIDAS DE VARIABILIDAD 5

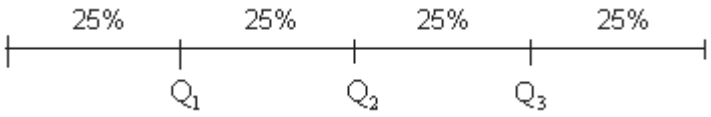
Universidad de El Salvador, Derechos Reservados 2013

MEDIDAS DE POSICIÓN

Las medidas de posición se usan para describir la posición de datos específicos con respecto al resto de los datos.

Los cuartiles: Q_k

Sean $x_1, x_2, \ldots, x_n$ un conjunto de n datos ordenados por orden de magnitud. El cuartil uno, Q_1, es aquel que deja a la izquierda ¼ (25%) de los datos y es menor que ¾ (75%) de los datos. El cuartil dos, Q_2, es la mediana. El tercer cuartil, Q_3, sobrepasa ¾ (75%) de los datos y es menor que ¼ (25%) de ellos.



Los cuartiles para datos no agrupados

Existen diferentes formas de calcular cuartiles, pero todas dan aproximadamente el mismo resultado. Un método cuando se calcula manualmente el k-ésimo cuartíl, Q_k, es el siguiente:

1. Ordenar los datos del más pequeño al más grande

2. Calcular $k(n/4)$

Se obtiene un entero ***w:***	*Se obtiene un Número mixto:* $W\ \frac{a}{b}$
3. La posición de Q_k es: $w + ½ = w.5$, es decir, Q_k está entre X_w y X_{w+1}	3. La posición de Q_k es: $w+1$
4. $Q_k = \frac{X_w + X_{w+1}}{2}$	4. $Q_k = X_{w+1}$

Universidad de El Salvador, Derechos Reservados 2013

Ejemplo 4.13.
Dados los valores 2, 5, 10, 3, 7, 13, 20 y 18, hallar Q_1 y Q_2.

Q_1: 1. Ordenando: 2, 3, 5, 7, 10, 13, 18, 20
2. Calcular $k(n/4)=1(8/4)=2$, de donde $w=2$
3. Q_1 está entre x_2 y x_3.
4. $Q_1 = \frac{x_2 + x_3}{2} = \frac{3+5}{2} = 4.0$.

Q_2: 1. 2, 3, 5, 7, 10, 13, 18, 20
2. Calcular $k(n/4)=2(8/4)=4$, de donde $w=4$
3. Q_2 está entre x_4 y x_5.
4. $Q_2 = \frac{x_4 + x_5}{2} = \frac{7+10}{2} = 8.5$

Ejemplo 4.14.
Dados los valores 9, 7, 6, 5 y 1, hallar Q_2 y Q_3.

Q_2: 1. 1, 5, 6, 7, 9
2. Calcular $k(n/4)=2(5/4)=2.5=2½$, de donde $w=2$
3. La posición de Q_2 es: $w+1= 3$
4. $Q_2 = x_3 = 6$

Q_3: 1. 1, 5, 6, 7, 9
2. Calcular $k(n/4)=3(5/4)=3.75$, de donde $w=3$
3. La posición de Q_3 es: $w+1=4$
4. $Q_3 = x_4 = 7$

Los cuartiles se usan con frecuencia en los datos de ventas y encuestas para dividir las poblaciones en grupos. Por ejemplo, se pueden utilizar para determinar el 25 por ciento de ingresos más altos en una población.

Los **cuartiles para datos agrupados** se calculan de manera semejante al cálculo de la mediana, donde $n/2$ se sustituye por $k(n/4)$. Así,

$$Q_k = l_i + c\left[\frac{k(n/4) - F_a}{f}\right]; k=1, 2, 3.$$

Universidad de El Salvador, Derechos Reservados 2013

Ejemplo 4.15. Hallar el cuartil tres de la siguiente distribución de frecuencias.

	x	f	F_a
0 - <10	5	2	2
10 - < 20	15	1	3
20 - < 30	25	5	8
30 -< 40	35	8	16
40 - < 50	45	4	20
		20	

Aquí, *k(n/4)*=3(20/4)=15, entonces ¿cuál es la primer frecuencia acumulada ascendente que contiene a 15?, *¡16!*, entonces la clase del cuartil tres es 30 - < 40, donde *li*=30, *Fa*=8 y *f*=8.

Así, al utilizar la fórmula se obtiene

$$Q_k = l_i + c\left[\frac{k(n/4) - F_a}{f}\right]$$

$$Q_3 = 30 + 10\left[\frac{15 - 8}{8}\right] = 38.8 \text{ años.}$$

Los percentiles o centiles: P_k ó C_k

Los valores que dividen al conjunto de datos en 100 partes iguales se llaman ***percentiles*** *o* ***centiles.***

Por ejemplo P_{90} para un conjunto de datos es el valor que deja debajo de él al 90% de ellos y arriba de él al 10%.

Los percentiles para datos no agrupados

El procedimiento es semejante al de los cuartiles.

Ejemplo 4.16.

Encontrar P_{70} en los siete valores 6, 10, 13, 19, 18, 16, y 15.

1. Ordenando: 6, 10, 13, 15, 16, 18, 19.
2. Calcular *k(n/100)*=70(7/100)=4.9, de donde *w*=4
3. La posición de P_{70} es: *w+1*=5
4. $P_{70} = x_5 = 16$

Esta medida permite establecer un umbral de aceptación. Por ejemplo, podría examinarse a los candidatos en un evento cuya calificación sea superior al nonagésimo percentil.

Universidad de El Salvador, Derechos Reservados 2013

Los **percentiles para datos agrupados** se calculan de manera semejante al cálculo de los cuartiles, donde *k*(*n/4)* se sustituye por *k(n/100).* Así,

$$P_k = l_i + c\left[\frac{k(n/100) - F_a}{f}\right]; k=1, 2,\ldots,99.$$

Ejemplo 4.17.

Considérese la tabla de distribución de frecuencias del ejemplo 4.15 y calcule el percentil 75.

Aquí, *k(n/100)*=*75*(20/100)=15, entonces ¿cuál es la primer frecuencia acumulada ascendente que contiene a 15?, *¡16!*, entonces la clase del percentil 75 es 30 - < 40, donde *li*=30, *Fa*=8 y *f*=8.

Así, al utilizar la fórmula se obtiene

$$P_k = l_i + c\left[\frac{k(n/100) - F_a}{f}\right]$$

$$Q_3 = 30 + 10\left[\frac{15 - 8}{8}\right] = 38.8 \text{ años.}$$

Universidad de El Salvador, Derechos Reservados 2013

MEDIDAS DE VARIABILIDAD

El grado en que los datos numéricos tienden a extenderse alrededor de un valor medio se le llama variación o dispersión de los datos.

Una medida de dispersión es importante por dos razones: primero, puede ser usada para mostrar el grado de variación entre los valores en los datos dados; y segundo, puede ser usada para suplementar un promedio para describir un grupo de datos o para comparar un grupo de datos con otro.

La desviación media: *D. M.*

*La **desviación media,** de una serie de n datos $x_1, x_2, \ldots, x_n$ viene definida por*

$$D.M. = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

donde $\bar{x}$ es la media aritmética de los valores y $|x_i - \bar{x}|$ es el valor absoluto de las desviaciones de las diferentes x_i de $\bar{x}$.

Ejemplo 4.18.
Hallar la desviación media de los valores 2, 3, 6, 8, 11.

La media es $\bar{x} = \frac{2+3+6+8+11}{5} = 6.0$. Entonces la *D. M.* es

$$D.M. = \frac{|2-6|+|3-6|+|6-6|+|8-6|+|11-6|}{5} = \frac{4+3+0+2+5}{5} = 2.8$$

La desviación media para datos agrupados

Aquí, $D.M. = \frac{\sum_{i=1}^{n} f_i |x_i - \bar{x}|}{n}$, donde x_i es la marca de clase y f_i la correspondiente frecuencia de clase.

Universidad de El Salvador, Derechos Reservados 2013

Ejemplo 4.19.
Hallar la *D. M.* a partir del ejemplo 4.2.

Calificación	f	x	$f\|x - \bar{x}\|$
3.70 - < 4.44	2	4.07	6.54
4.44 - < 5.18	2	4.81	5.06
5.18 - < 5.92	4	5.55	7.16
5.92 - < 6.66	6	6.29	6.30
6.66 - < 7.40	10	7.03	3.10
7.40 - < 8.14	10	7.77	4.30
8.14 - < 8.88	9	8.51	10.53
8.88 - < 9.62	7	9.25	13.37
Total	n=Σf=50		$\Sigma f\|x - \bar{x}\| =$ 56.36

$$D.M. = \frac{\sum_{i=1}^{n} f_i \left| x_i - \bar{x} \right|}{n} = \frac{56.36}{50} = 1.127.$$

La varianza: s^2

Si $x_1, x_2, \ldots, x_n$, es una muestra de n datos, entonces la ***varianza*** *muestral es*

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Las unidades de medición de la varianza muestral son iguales al cuadrado de las unidades con que se mide la variable. Una varianza grande indica bastante variación de datos. La varianza poblacional se denota por: σ^2.

Si un conjunto de datos constituye una población, o bien, si el objeto de resumir los datos es únicamente para describirlos, entonces se utiliza *n*. Si el propósito es sacar inferencias respecto a una población, entonces se debe sustituir el denominador, *n-1* por *n.*

$$\text{Así, } s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}, \quad s^2 = \frac{n\sum x^2 - \left(\sum x\right)^2}{n(n - 1)} \quad \text{ó} \quad s^2 = \frac{\sum x^2 - \left[\left(\sum x\right)^2 / n\right]}{n - 1}$$

Universidad de El Salvador, Derechos Reservados 2013

Ejemplo 4.20.
Hallar la varianza de los valores 63, 61, 65, 62, 61, 64, 60 y 66.

$$\sum x^2 = 31,532; \quad \sum x = 502.$$

$$s^2 = \frac{8(31,532)-(502)^2}{8(7)} = \frac{252,256-252,004}{56} = \frac{252}{56} = 4.5$$

Varianza para datos agrupados
La varianza para datos agrupados se calcula utilizando las fórmulas

$$s^2 = \frac{\sum f(x-\bar{x})^2}{n-1} \quad \text{o} \quad s^2 = \frac{\sum x^2 f - \left[\left(\sum xf\right)^2 / n\right]}{n-1}$$

Si s^2 se calcula sin fines inferenciales, o bien, si se calcula s^2 de una población, entonces se sustituye *n-1* por *n.*

Ejemplo 4.21.
Hallar la varianza de la siguiente distribución de frecuencias.

	x	f
0 - <10	5	2
10 - < 20	15	1
20 - < 30	25	5
30 -< 40	35	8
40 - < 50	45	4
		20

	x	f	xf	$x^2 f$
0 - <10	5	2	10	50
10 - < 20	15	1	15	225
20 - < 30	25	5	125	3,125
30 -< 40	35	8	280	9,800
40 - < 50	45	4	180	8,100
		20	610	21,300

Universidad de El Salvador, Derechos Reservados 2013

$$s^2 = \frac{\sum x^2 f - \left[\left(\sum xf\right)^2 / n\right]}{n-1} = \frac{21{,}300 - \left[(610)^2/20\right]}{19} = 141.8$$

La desviación estándar o típica: $s = \sqrt{s^2}$ **.**

*La **desviación estándar** es simplemente la raíz cuadrada positiva de la varianza, o sea*

$$s = \sqrt{s^2}$$

Ejemplo 4.22.
Halle la desviación típica en el ejemplo 4.20.

$$s^2 = 4.5, \quad s = \sqrt{s^2} = \sqrt{4.5} = 2.1$$

La desviación típica o estándar es una de las medidas de dispersión más importante. Es la que mejor mide cuanto se dispersan los datos con respecto a su media. Las unidades de la desviación estándar son las mismas que las de la medida.

El *coeficiente* de variación: *CV*

En ocasiones es deseable expresar la variación como fracción de la media. Para hacer esto se utiliza una medida adimensional de variación relativa, denominada **coeficiente de variación muestral.**

*El **coeficiente de variación muestral** es* $CV = \frac{s}{\bar{x}}$ *ó en forma porcentual*

$$CV = \frac{s}{\bar{x}} .100$$

El coeficiente de variación es útil cuando se compara la variabilidad de dos o más conjuntos de datos que difieren de modo considerable en la magnitud de las observaciones. Por ejemplo, el coeficiente de variación puede ser de utilidad al comparar la variabilidad del consumo de energía eléctrica por día en muestras tomadas en distintos conjuntos residenciales durante un determinado mes del año.

Universidad de El Salvador, Derechos Reservados 2013

Ejemplo 4.23.

Con un micrómetro, se realizan mediciones del diámetro de un balero, que tienen una media de 4.03 mm y una desviación estándar de 0.012 mm; con otro micrómetro se toman mediciones de la longitud de un tornillo, que tienen una media de 1.76 pulg y una desviación estándar de 0.0075 pulg ¿Cuáles de las medidas presenta una menor variabilidad?

El coeficiente de variación del balero y del tornillo son respectivamente

$$CV = \frac{0.012}{4.03} = 0.003 \text{ ó } 0.3\% \quad \text{y} \quad CV = \frac{0.0075}{1.76} = 0.004 \text{ ó } 0.4\%$$

Las medidas hechas con el primer micrómetro presentan una variabilidad relativamente menor que las efectuadas con el otro micrómetro.

Si el *CV* para un conjunto de datos es menor que *0.1* ó *10%*, entonces se dice que este conjunto de datos es homogéneo, es decir, hay muy poca variabilidad entre ellos y por lo tanto la media aritmética es altamente representativa de dichos datos.

GRADO DE REPRESENTATIVIDAD DE LA MEDIA

Si el *CV* porcentual es:

0% o menos de 10%, la media es altamente representativa.
10% o menos de 20%, la media es bastante representativa.
20% o menos de 30%, la media es representativa.
30% o menos de 40%, la media tiene representatividad dudosa.
40% o más, la media carece de representatividad.

Universidad de El Salvador, Derechos Reservados 2013